



University of Groningen

On the Reconstruction of a Weak Phase-Amplitude Object II. A new inversion theorem for finite Fourier transforms and the number of degrees of freedom of an image

Hoenders, B.J.; Ferwerda, H.A.

Published in:
Default journal

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1973

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Hoenders, B. J., & Ferwerda, H. A. (1973). On the Reconstruction of a Weak Phase-Amplitude Object II. A new inversion theorem for finite Fourier transforms and the number of degrees of freedom of an image. Default journal.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

On the Reconstruction of a Weak Phase-Amplitude Object

II. A new inversion theorem for finite Fourier transforms and the number of degrees of freedom of an image¹

By *B. J. Hoenders*² and *H. A. Ferwerda*

Technical Physical Laboratories, University of Groningen, The Netherlands

Received 16 October 1972

Abstract

In this article we derive a new inversion theorem for Fourier transforms of functions with finite support. With the aid of this theorem it shall be shown that the object can always be reconstructed with a prespecified tolerance from the values of the contrast at a sufficiently large set of points. These sampling points can always be chosen such that they lie within the visible part of the image plane. With the aid of a new sampling theorem we shall give an analysis of the concept of the number of degrees of freedom of an image, and show with a calculation similar to the one given by *Shannon* [2] that this number even might grow to infinity.

Inhalt

Über die Rekonstruktion eines schwachen Phasen-Amplituden-Objekts. II. Ein neues Inversionstheorem für endliche Fourier-Transformationen. Die Objektamplitude kann mit Hilfe dieses Theorems in vorgeschriebener Genauigkeit zurückgerechnet werden aus der Intensität in einer ausreichend großen Zahl von Punkten. Diese Samplingpunkte können stets im sichtbaren Teil des Bildfeldes liegen. Mit Hilfe des neuen Samplingtheorems analysieren wir die Anzahl der Freiheitsgrade im Bild, wir zeigen – ähnlich zu einer Rechnung von *Shannon* [2] –, daß diese Anzahl auch unendlich werden kann.

1. Introduction

In a previous article, *Hoenders* [1], it has been shown that a weak phase-amplitude object, imaged by an electron microscope with aberrations like coma and spherical aberration, can be reconstructed with a prespecified tolerance from the knowledge of the contrast in a finite part of the image

¹ Part I: *Optik* 35 (1972) 116.

² Part of a PhD-thesis submitted to the State University at Groningen, The Netherlands (June 1972).

plane. Noise was left out of consideration. However, the possibility to reconstruct a weak phase-amplitude object with an arbitrary accuracy seems to be in complete contradiction with a well-known concept in light optics, i.e. the concept of the number of degrees of freedom of an image, which concept applies just as well to electron optics. For, whichever formulation of this concept we take, be it the one given by *Shannon* [2], *Gabor* [3], or *Torvaldo di Francia* [4], in essence they all state that the image never contains enough information to reconstruct the object unambiguously.

Hence on one hand we have proven the possibility to reconstruct a weak phase-amplitude object with arbitrary precision and on the other hand the concept of the number of degrees of freedom of an image states the impossibility of such a reconstruction. In order to solve this contradiction we shall give a survey of several formulations of the concept of the number of degrees of freedom of an image together with the criticism which has already been raised against them. We shall also establish, in section 2, a new inversion theorem for Fourier transforms of functions with finite support like

$$g(\lambda) = \int_{-1}^{+1} e^{i\lambda y} h(y) dy. \quad (1.1)$$

The inversion of (1.1) i.e. the unambiguous determination of the function $h(y)$ up to a prespecified tolerance, if $g(\lambda)$ can be measured in some interval with a sufficiently small error, is especially relevant in optics. Referring to fig. 1 we specify the object in this paper by the wave function in a plane $z = z_0$ ("object plane") situated immediately behind the object. In this paper we shall be concerned with the determination of the wave function in a plane $z = z_0$ from the contrast in the image plane. So we deal with a two-dimensional reconstruction problem. The three-dimensional image reconstruction will be discussed in a subsequent paper. The finite Fourier transform arises as follows: If we consider an ideal³ optical system, the wave function in the back focal plane of the objective lens is the Fourier transform of the

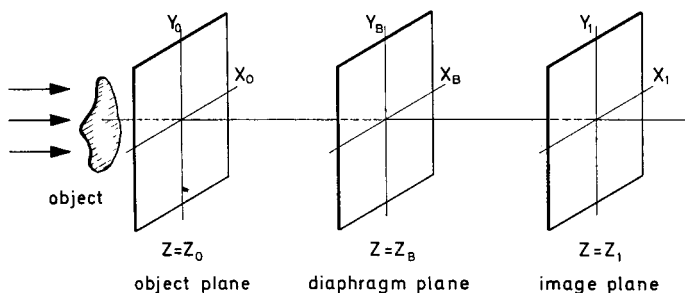


Fig. 1. General imaging process.

³ The case that aberrations like spherical aberration, coma and defocusing are present, has been treated in the previous article [1].

wave function in the object plane. In the back focal plane one has a real or effective aperture stop (diaphragm). The wave function in the Gaussian image plane is again the Fourier transform of the wave function in the back focal plane (see *Born and Wolf* [5], § 8.6.3). The wave function in the back focal plane has finite support (the wave function is supposed to be zero on the opaque part of the diaphragm), so the wave function in the image plane is an analytic function (see *Whittaker and Watson* [6], § 5.3.1). For practical reasons only a finite portion of the image plane is accessible to observation. We shall see in section 2 that this imposes no limitations on the accuracy of the reconstruction, provided noise may be neglected. It should further be emphasized that the wave function itself is not an observable quantity, only the absolute square of the wave function is. In the case we deal with a *weak* amplitude-phase object there is a linear relation between the contrast in the image plane and the amplitude attenuation and the phase shift (see formula (4.12) of [1]). As has been discussed in [1] the functions describing the amplitude attenuation (f) and phase shift (η) due to the object can be derived from the contrast of two electron micrographs taken at two different values of the defocussing. In [1], equations (4.23 a and b) the double finite Fourier transforms of f and η are expressed in terms of the contrast in the image plane. f and η can be calculated by inverting the double finite Fourier transforms.

Equation (1.1) states the problem as simply as possible. $g(\lambda)$ may be thought of as the contrast in the image plane while $h(y)$ stands for the phase shift or amplitude attenuation. $g(\lambda)$ is only known on a finite interval of the real axis (because the accessible part of the image plane is finite) up to a certain accuracy. The problem is how to calculate $h(y)$.

It shall be shown in section 2, corollary 1 of theorem 2, that the reconstruction of $h(y)$ can be performed uniquely up to a prespecified tolerance, if $g(\lambda)$ can only be measured with a sufficiently small uncertainty on a finite interval of the real axis.

This new inversion theorem is an alternative for the one given in appendix A of the previous article [1], and has the advantage that one only has to insert the measured value of $g(\lambda)$ at a finite set of sampling points in order to perform the inversion procedure.

Subsequently it shall be shown with the aid of a new sampling theorem that the whole concept of the number of degrees of freedom of an image lacks a rigorous foundation. For by using this new sampling theorem instead of the well-known Whittaker-Shannon sampling theorem one is led to the conclusion that the image contains an *infinite* number of degrees of freedom. It is therefore very likely that the whole concept of the degrees of freedom of an image only can get a sound basis if noise is taken into account.

2. A new sampling theorem for functions with finite band width and a new inversion theorem for finite Fourier transforms

Before we give an analysis of the concept of the degrees of freedom of an image we first derive some theorems which provide us with the necessary

mathematical background. The basic result of this section is a new sampling theorem which expresses a function $g(\lambda)$, which is the Fourier transform of a function with finite support in terms of the values $g(\lambda_j)$ of $g(\lambda)$ at a set of sampling points $\lambda = \lambda_j$. This set can always be chosen such that they all lie within the interval where $g(\lambda)$ can be measured. In order to do this we generalize a method due to *Filon* [7] (see also *Watson* [14], § 19.21) in order to expand a regularized δ -function denoted by $\delta_c(z - z_0)$ and defined by:

$$\delta_c(z - z_0) = \frac{\sin c(z - z_0)}{\pi(z - z_0)}, \quad (2.1)$$

into a set of discrete plane waves. Concerning the value of c see the note at the end of theorem 2. The expansion is formulated in the following theorem:

Theorem 1

Let the real numbers λ_j be defined by:

$$\lambda_j = a - \frac{1}{j}; \quad j = 1, 2, \dots \quad (2.2)$$

where the real number a has to be chosen such that $\lambda_j \neq 0$, if $j = 1, 2, \dots$. Then it is possible to expand the function $\delta_c(z - z_0)$, defined by:

$$\delta_c(z - z_0) = \frac{\sin c(z - z_0)}{\pi(z - z_0)}, \quad (2.3)$$

according to:

$$\delta_c(z - z_0) = \sum_{j=1}^p d_j(z, p) \exp(i\lambda_j z_0) + \varepsilon_p(z, z_0), \quad (2.4)$$

where

$$d_j(z, p) = -\frac{1}{2\pi} \int_{-c}^{+c} \frac{\exp(-ibz) \psi(b, p)}{(\lambda_j - b) \psi'(\lambda_j, p)} db, \quad (2.5)$$

$$\psi(\lambda, p) = \prod_{j=1}^p \left(1 - \frac{\lambda}{\lambda_j} \right), \quad (2.6)$$

and

$$\psi'(\lambda, p) = \frac{d}{d\lambda} \psi(\lambda, p). \quad (2.7)$$

If z_0 is restricted to a finite interval $|z_0| < G$, it is possible to determine for every positive number ε a number $P(\varepsilon)$, independent of z_0 and z for all real values of z , such that $|\varepsilon_p(z, z_0)| < \varepsilon$ for all $p > P(\varepsilon)$.

Proof

Consider the following contour integral, in which b is a real number and Q positive.

$$I(b, p, z_0) = \frac{1}{2\pi i} \oint_{|\lambda|=Q} \frac{\exp(i\lambda z_0)}{\psi(\lambda, p)(\lambda - b)} d\lambda; \quad (2.8)$$

we assume $b \neq \lambda_j$, $b \leq c < Q$, and $Q > \lambda_j$ ($j = 1, 2, \dots$).

The function $\{\psi(\lambda, p)\}^{-1}$ is analytic in the whole complex λ -plane except for a number of first order poles at $\lambda = \lambda_j$. Evaluating (2.8) using the theorem of residues we derive:

$$I(b, p, z_0) = \frac{\exp(ib z_0)}{\psi(b, p)} + \sum_{j=1}^p \frac{\exp(i\lambda_j z_0)}{(\lambda_j - b) \psi'(\lambda_j, p)} \quad (2.9)$$

valid because $b \neq \lambda_j$.

With (2.9) we derive from (2.8) using (2.5):

$$-\frac{1}{2\pi} \int_{-c}^{+c} \exp(-ibz) \psi(b, p) I(b, p, z_0) db = \delta_c(z - z_0) + \sum_{j=1}^p d_j(z, p) \exp(i\lambda_j z_0) \quad (2.10)$$

In the integration over b we have to exclude the point $b = \lambda_j$ because otherwise (2.9) is not valid. However, this exclusion will not change the value of the function $d_j(z, p)$, defined in (2.5), because we have excluded from the integration interval a set of measure zero where the integrand is finite. In the appendix we will prove the inequality:

$$\left| \frac{\psi(b, p)}{\psi(\lambda, p)} \right| < \left(\frac{1}{2}\right)^p, \quad (2.11)$$

valid if $|b| \leq c$, and if $|\lambda| = Q = 3\lambda_{\max} + \frac{2c\lambda_{\max}}{\lambda_{\min}}$, where λ_{\max} is the upper bound of $|\lambda_j|$ if $j = 1, 2, \dots$ and λ_{\min} is the lower bound of $|\lambda_j|$ if $j = 1, 2, \dots$. These numbers exist because $|\lambda_j|$ is a bounded sequence of real numbers if $j = 1, 2, \dots$. Furthermore λ_{\min}^{-1} exists because zero is not a limiting point of the sequence $|\lambda_j|$, by assumption.

From (2.11) we derive the following upper bound for the left hand side of (2.10):

$$\begin{aligned} & \left| \frac{1}{2\pi} \int_{-c}^{+c} \exp(-ibz) \psi(b, p) db \frac{1}{2\pi i} \oint_{|\lambda|=Q} \frac{\exp i\lambda z_0}{\psi(\lambda, p)(\lambda - b)} d\lambda \right| \\ & \leq \frac{1}{4\pi^2} \int_{-c}^{+c} |\psi(b, p)| db \oint_{|\lambda|=Q} \left| \frac{\exp i\lambda z_0}{\psi(\lambda, p)(\lambda - b)} \right| d|\lambda| \\ & \leq \frac{2c}{4\pi^2} \left(\frac{1}{2}\right)^p \exp(Qz_0) \frac{2\pi Q}{Q - c}. \end{aligned} \quad (2.12)$$

Hence, if z_0 belongs to a finite interval, such that $|z_0| < G$, there exists a

number $P(\varepsilon)$, independent of z_0 , such that the expression (2.12) is less than ε for every $p > P(\varepsilon)$. This completes the proof.

The choice of the sequence λ_j , as used in theorem 1, appears to be not unique. We could have chosen other sequences λ_j , e.g.

$$\begin{aligned} \text{a) } \lambda_j &= a - \frac{1}{(j+d)^n}; \text{ where } n \text{ is an arbitrary positive real number, and} \\ j &= 1, 2, \dots \\ \text{b) } \lambda_j &= a + \frac{d-a}{p}j; \quad j = 0, 1, 2, \dots, p \end{aligned} \quad (2.13)$$

where a and d are arbitrary positive numbers.

The points λ_j in (2.13b) are the sampling points if the interval between a and d in the complex plane is divided into p equal parts. This seems an effective way of sampling. In the following we shall use the definition (2.2) for the sampling points λ_j .

The next theorem deals with the inversion of a finite Fourier transform.

Theorem 2

Let $h(y)$ be a complex-valued function defined and of bounded variation in the interval $[-1, +1]$. Let $g(\lambda)$ (λ complex) be defined by

$$g(\lambda) = \int_{-1}^{+1} \exp(i\lambda y) h(y) dy. \quad (2.14)$$

Choose an arbitrary positive number ε . Then there exists a positive number $P(\varepsilon)$ and functions $d_j(y_0, p)$ defined in (2.5) such that

$$\sum_{j=1}^p d_j(y_0, p) g(\lambda_j) = h(y_0) + \varepsilon_p(y_0), \quad (2.15)$$

where $|\varepsilon_p(y_0)| < \varepsilon$ for all $p > P(\varepsilon)$ and $|y_0| \leq 1$.

Proof

From (2.4) and (2.14) one derives:

$$\sum_{j=1}^p d_j(y_0, p) g(\lambda_j) = \int_{-1}^{+1} \{\delta_c(y_0 - y) - \varepsilon_p(y_0, y)\} h(y) dy. \quad (2.16)$$

Because $h(y)$ is of bounded variation if $|y| \leq 1$, $h(y)$ is integrable on the interval $[-1, +1]$, and hence absolutely integrable on $[-1, +1]$. Let M be defined by:

$$\int_{-1}^{+1} |h(y)| dy = M. \quad (2.17)$$

According to theorem 1, there exists a positive number $P(\varepsilon)$ such that

$|\varepsilon_p(y_0, y)| < \frac{\varepsilon}{2M}$ for all $p > P(\varepsilon)$ if $|y| \leq 1$ and $|y_0| \leq 1$.

Using the theory of the Dirichlet integral (see Apostol [8], § 15.8), we infer the existence of a number c_1 such that⁴:

$$\left| \int_{-1}^{+1} \delta_{c_1}(y_0 - y) h(y) dy - h(y_0) \right| < \frac{1}{2} \varepsilon. \quad (2.18)$$

Hence, combination of (2.16), (2.17) and (2.18) leads to:

$$\left| \sum_{j=1}^p d_j(y_0, p) g(\lambda_j) - h(y_0) \right| < \frac{1}{2} \varepsilon + \frac{1}{2} \varepsilon = \varepsilon, \quad (2.19)$$

for all $p > P(\varepsilon)$, $|y_0| \leq 1$, and $c = c_1$ thus proving the theorem.

Corollary 1

If the uncertainty $|g(\lambda_j) - \tilde{g}(\lambda_j)|$ between the measured value $\tilde{g}(\lambda_j)$ at the sampling point $\lambda = \lambda_j$ and the "true value" $g(\lambda_j)$ is less than $\frac{\varepsilon}{2pN}$, where N equals the upper bound of $|d_j(y_0, p)|$ if $j = 1, \dots, p$ and $|y_0| \leq 1$, it follows from (2.19) that

$$\left| \sum_{j=1}^p d_j(y_0, p) \tilde{g}(\lambda_j) - h(y_0) \right| < 2\varepsilon;$$

for

$$\begin{aligned} & \left| \sum_{j=1}^p d_j(y_0, p) \{ \tilde{g}(\lambda_j) - g(\lambda_j) \} + \sum_{j=1}^p d_j(y_0, p) g(\lambda_j) - h(y_0) \right| \\ & < \sum_{j=1}^p |d_j(y_0, p)| | \tilde{g}(\lambda_j) - g(\lambda_j) | + \left| \sum_{j=1}^p d_j(y_0, p) g(\lambda_j) - h(y_0) \right| < 2\varepsilon. \end{aligned} \quad (2.20)$$

Hence, if the reconstruction of $h(y)$ has to be performed with a prespecified tolerance 2ε , it is sufficient to measure the values $g(\lambda_j)$ up to an uncertainty

$$\frac{\varepsilon}{2pN}.$$

⁴ The value of the number c_1 depends upon the particular function $h(y)$ which function is not known, however. Hence we have got the problem to determine a value of c_1 such that the function $h(y)$ can be reconstructed according to (2.19), which value only can be determined if $h(y)$ is known in advance. However, as will be pointed out in the next paper where we consider the threedimensional reconstruction of an object, if we have the a priori information that $h(y)$ belongs to a class of functions the total variation of which in the interval $|y| \leq 1$ is less than some number V , c_1 only depends upon V . Hence, we can choose a value of c_1 , such that (2.18) is valid, where the choice of c_1 only depends upon the a priori information that the total variation of the function $h(y)$ in the interval $|y| \leq 1$ is less than some number V .

Corollary 2

From (2.14) and (2.15) we now obtain:

$$g(\lambda) = \sum_{j=1}^p g(\lambda_j) \int_{-1}^{+1} \exp(i\lambda y) d_j(y, p) dy + \int_{-1}^{+1} \exp(i\lambda y) \varepsilon_p(y) dy \quad (2.21)$$

where $\left| \int_{-1}^{+1} \exp(i\lambda y) \varepsilon_p(y) dy \right| < 2\varepsilon$ for all $p > P(\varepsilon)$ and $-\infty < \lambda < +\infty$.

It is interesting to notice that $P(\varepsilon)$ is independent of λ !

(2.21) is a *sampling theorem* for $g(\lambda)$: $g(\lambda)$ is completely determined by its values in the points $\lambda = \lambda_j$. If we want to use formula (2.21) in practice we have to consider the influence of the inaccuracies in the measured values $\tilde{g}(\lambda_j)$ of the function $g(\lambda)$ at the sampling points λ_j .

Often the problem arises how the function $g(\lambda)$, which we know to be analytic in the whole complex λ plane from well-known theorems of function theory (see *Whittaker* and *Watson* [6], § 5.31) can be continued analytically outside the interval in which we know $g(\lambda)$ up to a given accuracy. *Wolter* [9] discusses this in the example of two analytic functions $f(z)$ and $g(z)$:

$$f(z) = \frac{1}{z+2}$$

$$g(z) = \frac{1}{z+2} + \frac{\varepsilon}{z+10^6}$$

(where ε is a small positive number),

which can hardly be distinguished on the interval $[-1, +1]$ on the real axis, but outside this interval $g(z)$ has a pole at $z = -10^6$ while $f(z)$ has not. Therefore, in a small interval around the point $z = -10^6$ $g(z)$ differs considerably from $f(z)$.

This example shows that a small uncertainty in the measurement of an analytic function in an interval, where this function is accessible to observation, can cause enormous uncertainties in the analytical continuation of the measured function outside that interval.

In theorem 3 it will be proved, that the image contrast can be determined with a prespecified tolerance in the whole image plane from the knowledge of this function in a finite number of sampling points, even if one makes a certain error in the measurement of the function.

It is clear that noise will determine the ultimate limit.

Theorem 3

Let $h(y)$ be a complex-valued function, defined and of bounded variation in the interval $[-1, +1]$. Let $g(\lambda)$ be defined by:

$$g(\lambda) = \int_{-1}^{+1} \exp(i\lambda y) h(y) dy. \quad (2.22)$$

If $\tilde{g}(\lambda_j)$ is the measured value of $g(\lambda)$ at the sampling point $\lambda = \lambda_j$ with error δ_j , which means that:

$$|g(\lambda_j) - \tilde{g}(\lambda_j)| < \delta_j,$$

we may calculate the function $g_{cp}(\lambda)$ from the p values $\tilde{g}(\lambda_j)$ according to:

$$g_{cp}(\lambda) = \sum_{j=1}^p \tilde{g}(\lambda_j) \int_{-1}^{+1} \exp(i\lambda y) d_j(y, p) dy \quad (2.23)$$

[compare (2.23) with (2.21)].

Let M_p be the maximum of $|d_j(y, p)|$ if $|y| \leq 1$, where $j = 1, \dots, p$. If ε is an arbitrary positive number, and if the errors δ_j fulfil the following condition:

$$\sum_{j=1}^p \delta_j < \frac{\varepsilon}{4 M_p} \quad (2.24)$$

then a number $P(\varepsilon)$ exists such that for every value of $p > P(\varepsilon)$

$$|g(\lambda) - g_{cp}(\lambda)| < \varepsilon \quad \text{for every real } \lambda. \quad (2.25)$$

Proof

From the definition of M_p we derive:

$$\left| \int_{-1}^{+1} e^{i\lambda y} d_j(y, p) dy \right| < 2 M_p; \quad (2.26)$$

valid for $j = 1, \dots, p$ and all real λ .

Let ε be an arbitrary positive number. From (2.21) we derive that there exists a number $P(\varepsilon)$ such that:

$$\left| g(\lambda) - \sum_{j=1}^p g(\lambda_j) \int_{-1}^{+1} \exp(i\lambda y) d_j(y, p) dy \right| < \frac{\varepsilon}{2} \quad (2.27)$$

for every fixed value of⁵ $p > P(\varepsilon)$ and all λ in $-\infty < \lambda < +\infty$.

From (2.24), (2.26) and (2.27) we then derive:

$$\begin{aligned} |g(\lambda) - g_{cp}(\lambda)| &= \left| g(\lambda) - \sum_{j=1}^p \tilde{g}(\lambda_j) \int_{-1}^{+1} \exp(i\lambda y) d_j(y, p) dy \right| \\ &\leq \left| g(\lambda) - \sum_{j=1}^p g(\lambda_j) \int_{-1}^{+1} \exp(i\lambda y) d_j(y, p) dy \right| + \\ &+ \left| \sum_{j=1}^p (g(\lambda_j) - \tilde{g}(\lambda_j)) \int_{-1}^{+1} \exp(i\lambda y) d_j(y, p) dy \right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

⁵ (2.27) is valid for every $p > P(\varepsilon)$. However, we do not consider the limit $p \rightarrow \infty$, because M_p might tend to infinity for $p \rightarrow \infty$. This would imply the value zero for every δ_j .

for every fixed $p > P(\varepsilon)$ and every λ in $+\infty < \lambda < +\infty$. This proves theorem 3.

3. Discussion of the concept "Degrees of freedom of an image"

3.1 Introduction

In physics, the minimum number of parameters which are necessary to specify a quantity is called the number of degrees of freedom of this quantity. In the present section we shall give a survey of the historical development of the concept of degrees of freedom of an image, together with the objections which have been raised against this concept. It will be shown, using the results of § 2, that even a more solid basis to some of the objections can be given, and that the reconstruction procedure proposed by *Wolter* [9] (see below) can be much simplified. Furthermore, we will show, using our new sampling theorem (2.21) that the conventional calculation of the number of degrees of freedom of an image becomes highly questionable.

3.2 Discussion

Shannon [2] has introduced the number of degrees of freedom for a band-limited signal. His discussion pertained to signals which are functions of *time*. In optics we also deal with band-limited signals. In that case the signal is the image, which is a function of the *spatial* coordinates in the image plane. The band limitation is due to the presence of an (effective) diaphragm in the optical system. Shannon's analysis can straightforwardly be extended to optical images. In theorem 5, we give the numerical value of the number of degrees of freedom and the assumptions under which it is derived. For reasons of brevity we restrict ourselves to one single spatial dimension of the object. The generalization to two dimensions is obvious.

Theorem 5 (*Shannon*)

Let the complex amplitude distribution to be imaged be given by the complex-valued function $\psi(x_0)$ (object wave function). Let the complex amplitude distribution in the image plane be given by $g(x_1)$ (image wave function). If the effective diaphragm has an aperture $-\frac{a}{2} \leq x_B \leq +\frac{a}{2}$, where x_B is the coordinate in the diaphragm plane and if the values of the coordinate x_1 in the image plane are restricted to $-\frac{w}{2} \leq x_1 \leq +\frac{w}{2}$ then the number of degrees of freedom of the image is $S = \frac{aw}{\lambda f}$, where λ is the wave length of the radiation and f is the back focal length of the objective.

Proof

It is well-known in optics (*Born and Wolf* [5], § 8.6.3) that a diaphragm in the back focal plane restricts the range of transmitted spatial frequencies. This implies that the image wave function $g(x_1)$ in the Gaussian image plane which is the Fourier transform of the complex amplitude distribution in the plane of the diaphragm, is a band-limited function of the spatial coordinate. Consequently, we can use the *Whittaker-Shannon* sampling theorem (see *Harris* [11]) which states:

$$g(x_1) = \sum_{l=-\infty}^{+\infty} \frac{\pi}{2} g\left(\frac{l\lambda f}{a}\right) \operatorname{sinc}\left\{\pi \frac{a}{f} \left(\frac{x_1}{\lambda} - \frac{l f}{a}\right)\right\}. \quad (3.1)$$

Formula (3.1) states the remarkable property that a band-limited function is completely determined from its values at an infinite discrete set of sampling

points $x_1 = \frac{l\lambda f}{a}$. Because the values of the coordinate x_1 are restricted to

$-\frac{w}{2} \leq x_1 \leq +\frac{w}{2}$, only the number of $S = \frac{aw}{\lambda f}$ sampling points lie within

this interval. Since it is furthermore assumed that the image wave function is "sufficiently well determined" by its values at these S sampling points, S is considered as the number of degrees of freedom of the image.

This reasoning cannot be made rigorous. The function $g(x_1)$, being a Fourier transform of a function with finite support, is an analytic function in the whole complex x_1 -plane (cf. *Whittaker and Watson* [6], § 5.31). This means that the values of $g(x_1)$ in the interval where this function can be measured, determine $g(x_1)$ in the whole complex x_1 -plane. The trouble is, however, that one has to continue $g(x_1)$ analytically which is a procedure that needs a very careful consideration. *Harris* [11] has proposed a method for performing this, using the sampling theorem (3.1); but he did not consider the influence on the analytical continuation of errors in the measurement of $g(x_1)$ in the

interval $\left[-\frac{a}{2}, +\frac{a}{2}\right]$. This influence might be dramatic, as shown in § 2,

where we gave an example of two analytic functions which could be hardly distinguished in some interval on the real axis, but had a totally different behaviour in other parts of the complex plane. Although it is not immediately clear how the value of $g(x_1)$ in the sampling points outside the interval

$\left[-\frac{w}{2}, +\frac{w}{2}\right]$ in the image plane can be calculated, the information concerning

the object wave function contained in these points is not completely lost. The significance of the number S is therefore not clear. The same kind of criticism can be raised against the formulation of the number of degrees of freedom of an image as given by *Gabor* [3]. This formulation, in *Gabor's* own words, is reproduced in theorem 6.

Theorem 6 (Gabor [3])

"Assume that the object area, large compared with the square of the wavelength, is limited by a black screen. Assume also that there is a similar limitation in the aperture plane at a great distance from the object plane (see fig. 2). Then, in the domain limited by these two black screens there exist S independent solutions of the wave equation $\nabla^2 u + k^2 u = 0$, that is to say, solutions with $u = 0$ immediately behind the black screens and S is given by the formula:

$$S = \lambda^{-2} \iiint dx dy d(\cos \alpha_x) d(\cos \alpha_y). \quad (3.2)$$

x and y are the coordinates in the object plane, and $\cos \alpha_x$ and $\cos \alpha_y$ are the direction cosines of the geometrical optical rays leaving the object plane. Any progressive wave through the object area and through the aperture can be expanded in terms of these S eigensolutions, with no more than S complex coefficients".

Some sort of a proof of *Gabor's* expansion theorem was obtained by *Miyamoto* [12]. He derived the following formula for S valid if $S \rightarrow \infty$:

$$S = \lambda^{-2} \iiint dx dy d(\cos \alpha_x) d(\cos \alpha_y) \{1 + O(\log |S|/S^{\frac{1}{2}})\}. \quad (3.3)$$

(3.3) clearly expresses that (3.2) only makes sense for large values of S and consequently the expansion theorem cannot be expected to hold if $S \approx 1$.

Wolter [10], [13] has criticized the statement of *Gabor's* concerning the meaning of S . To be more explicit, consider the expansion of $g(x_1)$ in terms of the eigenfunctions of the wave equation. *Gabor* assumes that this expansion contains only S terms. *Wolter* gives another interpretation. Only S coefficients in the expansion mentioned above are "appreciably" different from zero. The other ones are "very small". But, in accordance with our discussion of theorem 5 these neglected expansion coefficients are not lost or irrelevant. *Wolter's* point of view is corroborated by experiments which give much more information about the object than could be possible if S were an unavoidable upper limit for the information content of the image.

Wolter pointed out that the principal limit to the measurement of an observable quantity is the *Heisenberg* uncertainty relation. Up till now we have

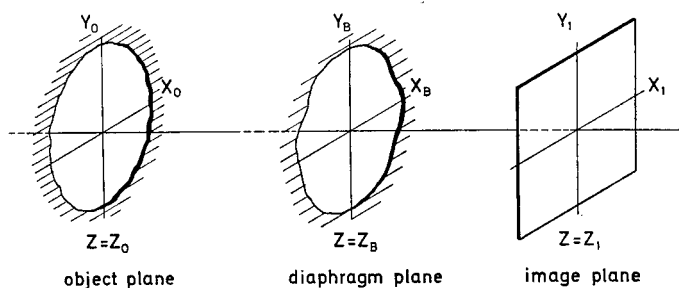


Fig. 2. Arrangement of the black screens in *Gabor's* expansion theorem.

not considered the influence of uncertainties made in the measurement of $g(x_1)$ in the aperture in the image plane. These uncertainties might influence the number of "detectable" expansion coefficients which is probably roughly equal to S . *Wolter* [9] has shown that even in the presence of errors there does not exist a number with the interpretation given to S . This is immediately understood from the following theorem 7:

Theorem 7 (Wolter [9], [10])

Let $\psi(x_0)$ be the wave function in the object plane, imaged by an objective, and let d be the width of the aperture in the object plane. Suppose $g(x_1)$ to be the image wave function, and let a be the width of the aperture of the diaphragm in the back focal plane of the objective. According to elementary optics (see *Born and Wolf* [5], § 8.6.3) there exists between $g(x_1)$ and $\psi(x_0)$ the following relation:

$$g(x_1) = \int_{-\frac{a}{2}}^{+\frac{a}{2}} \exp(ik \frac{x_1}{D} x_B) dx_B \int_{-\frac{d}{2}}^{+\frac{d}{2}} \exp(ik \frac{x_B}{f} x_0) \psi(x_0) dx_0$$

where f is the focal distance and D is the distance between the image and diaphragm plane.

For every positive ε there exists an interval $[-G(\varepsilon), +G(\varepsilon)]$, where the values of the image wave function have to be known within an accuracy $\delta(\varepsilon)$ such that from these values one may construct an approximating function $\psi_a(x)$ for $\psi(x)$ with the property that

$$\int_{-1}^{+1} |\psi_a(x) - \psi(x)|^2 dx < \varepsilon.$$

The application of theorem 7 may encounter some problems, for in order to calculate the function $\psi_a(x)$ one has to know the image wave function in an interval, the length of which depends on the tolerance of the reconstruction. But the part of the image plane where $g(x_1)$ is accessible to observation is finite. Therefore it can happen that the interval $[-G(\varepsilon), +G(\varepsilon)]$ extends outside the interval where $g(x_1)$ can be measured. Theorem 7 is for this reason only of limited value unless there is some recipe to continue analytically the measured wave function $g(x_1)$ (which is given in theorem 3). Theorem 3 tells therefore that S can certainly *not* be an upper limit to the information contained in the image: it is possible to reconstruct the object wave function to any degree of accuracy provided the errors in the measurement of the image contrast are sufficiently small. The sampling procedure discussed in theorem 3 uses an infinity of sampling points all lying in the "observable part" of the image plane. One might be tempted to say that the image contains an infinite number of degrees of freedom. The considerations given above show that the concept of "number of degrees of freedom" lacks a rigorous foundation, because the number seems to depend on the particular sampling pro-

cedure used (*Whittaker – Shannon* vs theorem 3). It should be clear that a discussion on the meaning of “number of degrees of freedom” only makes sense if the noise is taken into account.

Appendix

In this appendix we shall prove the inequality (2.11).

Theorem:

Let $\psi(b, p)$ be defined by:

$$\psi(b, p) = \prod_{j=1}^p \left(1 - \frac{b}{\lambda_j}\right), \quad (\text{A.1})$$

where the numbers λ_j are defined in (2.2).

Let λ_{\max} be the upper bound of $|\lambda_j|$ if $\lambda_j = 1, 2, \dots$ and λ_{\min} the lower bound of $|\lambda_j|$ if $j = 1, 2, \dots$. The sequence $\{\lambda_j\}$ is chosen such that λ_{\max} and λ_{\min} exist and are different from zero. Let c be an arbitrarily chosen positive number. If b is a real number chosen such that $|b| \leq c$ then the inequality

$$\left| \frac{\psi(b, p)}{\psi(\lambda, p)} \right| < \left(\frac{1}{2}\right)^p \quad (\text{A.2})$$

holds for all values of λ satisfying

$$|\lambda| = 3\lambda_{\max} + \frac{2c\lambda_{\max}}{\lambda_{\min}}. \quad (\text{A.3})$$

Proof:

From A.1 we derive:

$$\left| \frac{\psi(b, p)}{\psi(\lambda, p)} \right| \leq \frac{\prod_{j=1}^p \left(1 + \left|\frac{b}{\lambda_j}\right|\right)}{\prod_{j=1}^p \left(1 - \left|\frac{\lambda}{\lambda_j}\right|\right)} < \left(\frac{1 + \frac{c}{\lambda_{\min}}}{\frac{|\lambda|}{\lambda_{\max}} - 1}\right)^p \quad (\text{A.4})$$

Using (A.3) we derive from (A.4):

$$\left| \frac{\psi(b, p)}{\psi(\lambda, p)} \right| < \left(\frac{1 + \frac{c}{\lambda_{\min}}}{\frac{|\lambda|}{\lambda_{\max}} - 1}\right)^p < \left(\frac{1}{2}\right)^p.$$

References

- [1] *B. J. Hoenders*, *Optik* 35 (1972) 116.
- [2] *C. Shannon*, *The mathematical theory of communication*. University of Illinois Press, Urbana (1949).

- [3] *D. Gabor*, Proc. of the Symp. on Astronomical Optics. North-Holland Publishing Company, Amsterdam, p. 17 (1956).
- [4] *G. Toraldo di Francia*, J. Opt. Soc. Am. 59 (1969) 799.
- [5] *M. Born* and *E. Wolf*, Principles of Optics. Pergamon Press (1965).
- [6] *E. T. Whittaker* and *G. N. Watson*, A course of modern analysis. Cambridge University Press (1964).
- [7] *L. N. G. Filon*, Proc. London Math. Soc. (2) 4 (1906) 396.
- [8] *T. M. Apostol*, Mathematical Analysis, Addison-Wesley (1965).
- [9] *H. Wolter*, A. E. Ü. 13 (1959) 393.
- [10] *H. Wolter* in Progress in Optics, Vol. 1 (E. Wolf editor). North-Holland Publishing Company, Amsterdam 1961.
- [11] *J. L. Harris*, J. Opt. Soc. Am. 54 (1964) 931.
- [12] *K. Miyamoto*, J. Opt. Soc. Am. 50 (1960) 856.
- [13] *H. Wolter*, Physica 24 (1958) 457.
- [14] *G. N. Watson*, A treatise on the theory of Bessel functions. Cambridge University Press (1966).